Outbreak, Surveillance, Investigation & Response (OSIR) Journal
Field Epidemiology Training Program, Bureau of Epidemiology
Department of Disease Control, Ministry of Public Health, Thailand
Tel: +6625901734-5, Fax: +6625918581, Email: osireditor@osirjournal.net, http://www.osirjournal.net

Grammar of Science: To Boost "Odds" to Reduce "Risks" and to Avoid "Hazards"

Jaranit Kaewkungwal*

Mahidol University, Thailand

* Corresponding author, email address: jaranit.kae@mahidol.ac.th

A patient gets confused by listening to his doctors. Doctor A told him that "Knowing your risk factors for stroke is the first step in preventing a stroke. Risk factors that you can change or treat included, for examples, high blood pressure, smoking, diabetes, high cholesterol, physical inactivity and obesity, and sleep apnea. But the risk factors that you can't control are increasing age, gender, heredity and race."[1] Doctor B told him that "Nothing will help you prevent a stroke more than quitting smoking. Other important ways to lower your odds of having a stroke include lose weight, drink less alcohol, consume less sodium (salt), eat a healthy diet and spend less time in front of screens and more time walking."[2] Doctor C, warned him that "For people who were admitted to a hospital at the time of their index stroke and received the treatment in time, the chance of stroke recurrence was reduced by 16%. Based on their hazard ratios, factors associated with stroke recurrence include having comorbid conditions, both diabetes and urinary incontinence, and other cardiac conditions."[3]

So what are "Risk", "Odds" and "Hazard"?

By dictionary definition, "risk" is the possibility of loss or injury[4]. In epidemiology, however, "risk" is defined as "incidence". Statisticians further define "incidence" as "chance" or "probability" of developing outcome of interest no matter good or bad (e.g., disease, cure or die). Incidence means the occurrence of new outcomes in the study population over a specified period of time, and it also means the number of new cases per unit of population.[5,6] Thus, we can say that there are two types of incidence that are commonly used: incidence proportion and incidence rate. Incidence proportion or cumulative incidence is the proportion of a population that does not have the outcome (simply called disease-free population) and then some subsequently develop the outcome during a specified period of time. Basic statistical formula for Incidence proportion is number of new cases (numerator) divided by the total number of population (denominator); thus, a risk is a proportion. Incidence proportion does not take into consideration about time-at-risk (follow-up time from the starting point and still disease free, but the person is at risk of having the outcome).[7,8]

In a research study, if your question is about how many outcomes occur in the total population within a unit of time (e.g., per day, month, year), that means we want to know incidence rate or person-time rate. A person-time is an epidemiologic jargon and generally calculated from a total time of all people in a study contributed until they reach the "endpoints" (i.e., having the outcome of interest) or are "censored" (i.e., not having the outcome due to lost to follow-up or reaching the end of the study period). Basic statistical formula for Incidence rate is number of new cases (numerator) divided by the total time-at-risk of population (denominator). Thus, an incidence rate reflects how quickly disease occurs in a population.[6-8]

Odds can be defined as the risk (or probability) of an outcome occurring over the risk (or probability) of an outcome not occurring.[6] For example, if we follow 100 smoking people in a community for five years (each person contributes five years of follow-up time) and 10 of them eventually develop stroke at the end of 5 years. We now can say that among smokers the risk of having stoke is (10/100) 0.1 or 10%, the rate of having stroke is (10/500) 0.02 or 2% per year, and the odds of having stroke is 0.1/0.9 or 0.11:1.

By dictionary definition, a "hazard" is a source of harm or danger; where "danger" is exposure or liability to injury, pain, harm, or loss[9,10]. From this definition, hazard is danger, and risk is the probability of encountering the danger. However, in epidemiology, similar idea but not exactly the same as incident rate, the term "hazard" refers to the probability that a person has been followed and then develops an outcome or reaches the endpoint at time t[11,12]. We can say that hazard is the probability of an outcome occurrence of an individual, based on his/her "time-to-event" (so-called "survival time"); thus, hazard represents the instantaneous event rate for an individual who has already survived to the time "t"[11]. For examples, we can calculate a hazard of a diabetes patient to develop second episode of stroke after he has been followed from his first stroke.

Risk Comparisons – "Odds Ratio", "Risk Ratio" and "Rate Ratio"

Now we want to compare risks among those who have different exposures, which means that we want to assess a measure of association or relationship between exposure and outcomes among the two groups. Exposure is a generic epidemiologic term while it could be personal characteristics (e.g., gender, age, occupation, smoking), genetic/biologic characteristics (e.g., genotyping, immune status), acquired characteristics (e.g., disease status), or environmental characteristics (e.g., residential). Common measures of association include risk ratio (relative risk), rate ratio and odds ratio.[6,13]

A risk ratio or relative risk (RR) compares the risk of having the outcome of the two exposure groups. Basically, RR is calculated by dividing the risk (or incidence proportion) of one group against the risk in another group (baseline or reference group). A rate ratio (also abbreviated as RR) compares the incidence rates or person-time rates of the two groups. Odds ratio (OR) is another measure of association, comparing the odds of an outcome occurring in one group by the odds of the same outcome in another group.[6,8,12,13] As an example shown in figure 1, in a clinical trial, the AIDS patients with an initial episode of PCP (*Pneumocystis carinii* pneumonia) were randomly allocated to receive treatment A or B. Patients in each group were followed up, and some of them had PCP relapse. However, they were not all "relapsed" (reaching the endpoint) or "not relapsed" (being censored) at the same time. For example, patient obs#1 were followed and had relapse (pcp=1) at 11.9 months, while patient obs#2 were followed 11.6 months and not relapsed (pcp=0). The researchers then can compare the two treatments regarding the risk of having PCP relapse by calculating RR (risk/rate ratios) or OR as shown in figure 2.

	obs	trt	trtno	pcp	pdate
1.	1	B	0	1	11.9
2.	2	B	0	0	11.6
3.	3	A	1	0	12.8
4.	4	A	1	0	7.3
5.	5	B	0	1	4.5
6.	6	B	0	0	18.1
7.	7	A	1	0	14.7
8.	8	B	0	0	24
9.	9	A	1	0	16.2
10.	10	A	1	0	26.6

Figure 1. Example of raw data of a clinical trial to compare risk of relapse between two treatments

trtno	relapse (pcp: 1=yes,0=no 0	1	Total	Total_time (months)
0	120 77.42	35 22.58	155 100.00	2073.7
1	140 90.91	14 9.09	154 100.00	2379.6
Total	260 84.14	49 15.86	309 100.00	4453.3

Odds ratio = Odds trtnoA / Odds trtnoB
= [(Probability Yes : Probability No) trtnoA] / [(Probability Yes : Probability No) trtno B]
= [(14/154)/(140/154)] / [(35/155)/(120/155)]
= 0.343

Risk ratio = Risk trtno A / Risk trtnoB
= Incidence Proportion trtnoA / Incidence Proportion trtnoB
= Probability Yes trtnoA / Probability Yes trtnoB
= (14/154) / (35/155)
= 0.403

Rate ratio = Rate trtnoA / Rate trtno B
= Incidence Rate trtnoA / Incidence Rate trtno B
= (14/2379.6) / (35/2073.7)
= 0.349

Figure 2. Basic statistics for comparing risk of relapse between two treatments

By statistical formula, we will find that OR approximates risk ratio when the outcomes are rarely happened. OR cannot be used to estimate rate ratio because the denominator of the rate is time-at-risk. When should we present odds ratio or risk ratio? If the outcome is incidence, we can present either risk ratio or odds ratio; if not, we have to present OR[6,7,8]. There is a recommendation that no matter we select to present risk ratios and OR, we should give information about the frequencies of the outcome and the exposure risk factor[7].

Risk Comparisons – "Hazard Ratio"

As previously mentioned, time-to-event or survival time is the expected duration of time until one or more events happen. Although it is called survival time, but the event or endpoint does not have to always be "dead"; the researchers may want to study time from date of drug initiation until date the patient is cured. Analysis of time-to-event takes into consideration for both cases that have complete time from the starting point to reaching the endpoint and cases that have time from the starting point until they are censored. Censoring that is random and non-informative is usually required in order to avoid bias in a time-to-event analysis; thus, the analysis will correctly incorporate information from both censored and uncensored observations[14,15].

Based on the time-to-event and the event status (endpoint or censored), we can estimate two functions that are dependent on time, the survival and hazard functions.[14] Both functions describe the distribution of event times. The survival function gives, for every time, the probability of surviving (or not reaching the outcome) up to that time. On the opposite, the hazard function gives the potential that the outcome event will occur, per time unit, given that an individual has survived (or not yet having the outcome) up to the specified time[14]. Based on the example of a clinical trial among PCP patients who were randomly allocated to treatment A or B, each patient had different follow-up "time" in the study (Figure 3). Some were "relapsed" (so-called "failure" cases) and some were "not relapsed" (so-called "censored" or "net loss" cases) at different follow-up times. For example, among 155 patients in treatment A (trtno=0) group at the beginning, there was one relapsed case and none loss (or censored) at the time of 0.2 month; thus, there were 154 patients at the beginning of next time period and another one relapsed and none censored at the next time period of 1.1 month, and so on. From those events throughout each time period, we can calculate survival function (probability of "not relapse" over time) and hazard function (probability of "relapse" over time) as shown in figure 3.

Time	Beg. Total	Relapse	Net Lost	Prob Relapse	Prob Not relapse	Survivor Function		Hazard Function		Cummulative Hazard
trtno=0										
0.2	155	1	0	0.00645	0.99355	0.99355	*(0.99355 x 1)*	0.00645	*(1 - 0.99355) / 1*	0.00645
1.1	154	1	0	0.00649	0.99351	0.98710	*(0.99351 x 0.99355)*	0.00649	*(0.99355 - 0.98710) / (0.99355)*	0.01294
1.2	153	1	0	0.00654	0.99346	0.98065	*(0.99346 x 0.98710)*	0.00654	*(0.98710 - 0.98065) / (0.98710)*	0.01948
1.3	152	0	1	0.00000	1.00000	0.98065	*(1.00000 x 0.98065)*	0.00000	*(0.98765 - 0.98765) / (0.98765)*	0.01948
1.4	151	2	0	0.01325	0.98675	0.96766	*(0.98675 x 0.98065)*	0.01325	*(0.98765 - 0.96766) / (0.98765)*	0.03272
1.5	149	0	1	0.00000	1.00000	0.96766	*(1.00000 x 0.96766)*	0.00000	*(0.96766 - 0.96766) / (0.06766)*	0.03272
:	:	:	:	:	:	:	:	:	:	:
trtno=1										
0.1	154	1	1	0.00649	0.99351	0.99351	*(0.99351 x 1)*	0.00649	*(1 - 0.99351) / 1*	0.00649
0.4	152	1	0	0.00658	0.99342	0.98697	*(0.99342 x 0.99351)*	0.00658	*(0.99351 - 0.98697) / (0.99351)*	0.01307
0.6	151	1	0	0.00662	0.99338	0.98044	*(0.99338 x 0.98697)*	0.00662	*(0.98697 - 0.98044) / (0.98697)*	0.01969
2.9	150	0	1	0.00000	1.00000	0.98044	*(1.00000 x 0.98044)*	0.00000	*(0.98044 - 0.98044) / (0.98044)*	0.01969
5.2	149	0	1	0.00000	1.00000	0.98044	*(1.00000 x 0.98044)*	0.00000	*(0.98044 - 0.98044) / (0.98044)*	0.01969
5.5	148	0	1	0.00000	1.00000	0.98044	*(1.00000 x 0.98044)*	0.00000	*(0.98044 - 0.98044) / (0.98044)*	0.01969
:	:	:	:	:	:	:	:	:	:	:

Beg Total = Number of cases at the beginning of time period
Relapse = Number of relapse cases at the time period
Net lost = Number of cases who lost to follow-up or exit the study at the time period

Prob relapse = Probability of relapse (failure) = Relapse / Beg_total
Prob Not relapse = Probability of not relapse (survival) = 1 – Prob Relapse

Survival Function (*St*) = Probability of not relapse at the end of that period
= Prob Not relapse at time t x Prob not relapse at the end of time t-1

Hazard Function (*Hzt*) = Probability of relapse over time at time t ,
given that the person not relapse at time t-1
= *St* that changes at the end of time t from time t-1,
given that the person not relapse at time t-1
= *(St – St-1) / St-1*

Figure 3. Examples of survival function and hazard function of the two treatment groups

From the nonparametric estimators of the survival function (Figure 3), we usually present survival probabilities as a function over time using the Kaplan Meier graph as shown in figure 4. When we compare chance of reaching the outcome over time (hazard function) between two groups with different exposures (e.g., Treatment A-B, smoking Y-N), we will get "hazard ratio" (HR). Thus, we can say that HR is a measure of relative risk over time in circumstances where we are interested not only in the total number of events, but in their timing as well [8,14,15].

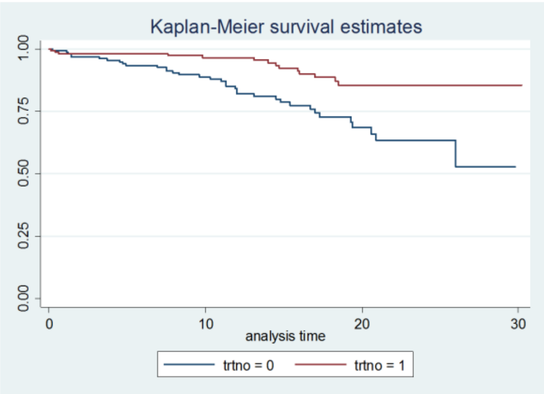

Figure 4. Kaplan-Meier survival function curve of two treatments

What can the Doctor Tell the Patient about "Odds", "Risks" or "Hazards"?

So we can calculate the OR/RR/HR from the study samples, but does it represent the "true" risk in population? Can you recommend your patient to boost or reduce odds (chance of having the outcome over not having the outcome), risk (chance of having the outcome), or hazard (chance of having the outcome over time)? Regression models that give you OR (logistic regression model), RR (Poisson regression model) and HR (Cox's proportional hazard model) usually provide hypothesis testing of the OR/RR/HR with p-value estimate.[6,7,14] In literature, sometimes they do not present p-value, but show OR/RR/HR with its 95% CI. Remember our definition of 95% CI in previous article: it represents the estimates of the true value in the population. For this statistics, if the 95% CI of the estimate does not include 1, we will say that such factor is statistically significant. For example, if OR of stroke between smoking versus not smoking groups is 4.5 (95% CI = 3.0-7.6), then we can say that the odds to have stroke were statistically significant different (increase) among smokers compared to non-smokers. If RR of stroke between male versus female is 2.8 (95% CI = 0.8-3.7), then we can say that the risk to have stroke were not statistically significant different between male and female. If HR of stroke between treatment A versus treatment B is 0.25 (95% CI = 0.2-0.5), then we can say that the risk to have stroke were statistically significant different (reduce) if the patients get treatment A compared to those who get treatment B. Note that when OR/RR/HR is 1, it means no statistically difference between comparison groups; when it is more than 1, that means one group has higher risk than its counterpart group (baseline/reference group); and when it is less than 1, that means one group has lower risk (protective) than its counterpart group. If the study is a clinical trial, we can also calculate "efficacy" of the treatment, technically called "prevented fraction among the exposed" from RR/HR; the formula is "Efficacy = 1-RR or 1-HR". For example, when HR of stroke between treatment A versus treatment B is 0.25 (95% CI = 0.15-0.45), then we can say that the efficacy of treatment A compared to treatment B is 75% (55-85%)[6,11].

Now the patient understands the terms "risks", "odds" and "hazards" that Doctor A, Doctor B and Doctor C are trying to tell him!

Suggested Citation

Kaewkungwal J. Grammar of science: to boost "odds" to reduce "risks" and to avoid "hazards". OSIR. 2018 Sep;11(3):22-6.

References

1. American Stroke Association. Let's talk about risk factors for stroke. 2017 [cited 2018 Sep 4]. <https://www.strokeassociation.org/idc/groups/stroke-public/@wcm/@hcm/documents/downloadable/ucm_309713.pdf>.
2. Harvard Health Publishing. How to lower your stroke risk. 2013 Aug [cited 2018 Sep 4]. <https://www.health.harvard.edu/heart-health/how-to-lower-your-stroke-risk>.
3. Lee AH, Somerford PJ, Yau KKW. Risk factors for ischaemic stroke recurrence after hospitalization. Med J Aust. 2004;181(5): 244-6.
4. Merriam-Webster. Risk [cited 2018 Sep 6], <http://www.merriam-webster.com/dictionary/risk>.
5. Cole SR, Hudgens MG, Brookhart MA, Westreich D. Risk. Am J Epidemiol. 2015;181(4):246-50.
6. Centers for Disease Control and Prevention. Principles of epidemiology in public health

practice: an introduction to applied epidemiology and biostatistics. 3rd ed. 2012 May [cited 2018 Sep 6]. <https://www.cdc.gov/ophss/csels/dsepd/ss1978/SS1978.pdf>.

7. Cummings P. The Relative merits of risk ratios and odds ratios. Arch Pediatr Adolesc Med. 2009 May;163(5):438-45.

8. Scott I. Interpreting risks and ratios in therapy trials. Aust Prescr. 2008;31:12-6.

9. Merriam-Webster. Hazard [cited 2018 Sep 6]. <http://www.merriam-webster.com/dictionary/hazard>.

10. Merriam-Webster. Danger [cited 2018 Sep 6]. <http://www.merriam-webster.com/dictionary/dange>.

11. Brody T. Clinical trials: study design, endpoints and biomarkers, drug safety, and FDA and ICH guidelines. 2nd ed. New York: Academic Press; 2016.

12. Stare J. Odds ratio, hazard ratio and relative risk. Metodoloˇski zvezki. 2016;13(1):59-67.

13. Jewell NP. Risk comparisons. Am J Ophthalmol. 2009;148(4):484-6.

14. Cornell Statistical Consultant Unit. Cornell University. What is Survival Analysis? [cited 2018 Sep 6]. <https://www.cscu.cornell.edu/news/statnews/stnews78.pdf>.

15. Hernán MA. The hazards of hazard ratios. Epidemiology. 2010;21(1):13-5.